34

Fiber-Optic Communications and Networks

MARTIN MAIER

SUMMARY

Optical technologies can be found at all network hierarchy levels of today's network infrastructures, ranging from fiber-to-the-home broadband access networks to high-speed backbone networks. In this chapter, we discuss recent progress on fiber-optic communications and networks and review the most important optical switching techniques that help improve the bandwidth efficiency and decrease the cost and power consumption of next-generation optical networks with subwavelength switching capabilities. Finally, we discuss remaining open issues and elaborate on future research challenges and opportunities.

INTRODUCTION

Optical fiber provides an unprecedented bandwidth potential that is far in excess of any other known transmission medium. A single strand of fiber offers a total bandwidth of 25,000 GHz. Optical fiber is commonly recognized as an excellent transmission medium owing to its advantageous properties, such as low attenuation, huge bandwidth, immunity against electromagnetic interference, longevity, and low maintenance costs. Because of their unique properties, optical fibers have been widely deployed to realize high-speed enterprise and service provider transmission links and networks [1].

Optical fiber links may carry either a single wavelength channel or multiple wavelength channels by means of wavelength division multiplexing (WDM). The advent of Erbium-doped fiber amplifiers (EDFAs) was the key to the commercial adoption of WDM links in today's network infrastructure. WDM links offer huge amounts of capacity and dramatically improve the unregenerated bandwidth distance product in a cost-effective manner. WDM links are widely deployed in metropolitan and wide area networks and are clearly one of the major success stories of fiber-optic communications. Optical WDM networks are primarily implemented in the form of simple ring topologies, particularly in metro networks, with a small amount of mesh topologies being deployed in long-haul networks [2].

Current optical metro and wide area networks offer large bandwidth pipes where a single fiber can carry tens or even hundreds of wavelength channels, each operating at a bit rate of 10 Gb/s or higher. Given these vast amounts of available bandwidth, one of the major design criteria of the current optical WDM backbone networks has been not to maximize the utilization of bandwidth resources but rather to simplify network operation and reduce capital expenditures (CAPEX) and operational expenditures (OPEX). Toward these goals, a few enabling optical technologies were successfully deployed, for example, reconfigurable optical add-drop multiplexer (ROADM), optical cross-connect (OXC), and tunable laser. Most current operational optical WDM backbone networks deploy circuit switching at the wavelength granularity. The resultant point-to-point wavelength channels are often referred to as *lightpaths* (or *light-trees* in the case of point-to-multipoint wavelength channels). At present, there is a strong worldwide push toward bringing fiber optics closer to individual homes and businesses. Fiber-to-the-home/business (FTTH/B) networks are poised to become the next major success story for fiber-optic communications [2,3].

Due to the ever-increasing speeds of access networks, the bandwidth bottleneck will move from the first/last mile toward metropolitan and wide area networks. To provide higher bandwidth efficiency, optical metro and core wavelength-switching networks could be required to resort to more efficient switching techniques at the sub-wavelength granularity in the near to mid term. Future optical networks may deploy a variety of optical switching techniques across the range of switching granularity, ranging from switching entire fibers to switching individual packets on each wavelength channel [4].

RECONFIGURABLE ADD–DROP MULTIPLEXER AND OPTICAL CROSS-CONNECT

Wavelength-switching *all-optical networks* (AONs) provide end-to-end optical paths by deploying all-optical node structures, which allow the optical signal to stay partly in the optical domain. Such all-optical nodes are also called OOO nodes to emphasize the fact that they do not perform costly optical–electrical–optical (OEO) conversion of all wavelength channels and in-transit traffic stays in the optical domain. OOO nodes come in different flavors, as explained in the following.

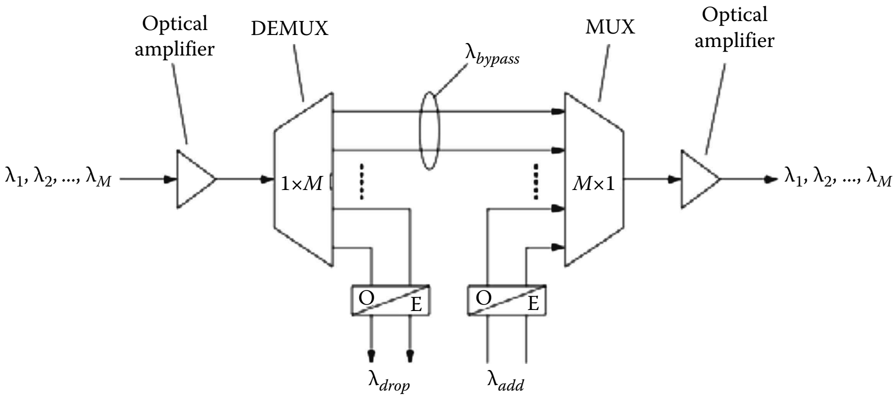

FIG. 34.1
Optical add–drop multiplexer (OADM).

Optical Add–Drop Multiplexer

Figure 34.1 shows the basic schematic of an *optical add–drop multiplexer* (OADM) with a single input/output fiber link that carries M different wavelength channels. At the input fiber, the incoming optical signal comprising a total of M wavelengths $\lambda_1, \lambda_2, \ldots, \lambda_M$ is pre-amplified by means of an optical amplifier. A good choice for an optical amplifier is the EDFA since a single EDFA is able to amplify multiple WDM wavelength channels simultaneously. After optical pre-amplification, the WDM wavelength comb signal is partitioned into its M separate wavelengths by using a $1 \times M$ wavelength demultiplexer (DEMUX). In general, some bypass wavelengths λ_{bypass} remain in the optical domain and are thus able to optically bypass the local node. The remaining wavelengths λ_{drop} are dropped by means of optical–electrical (OE) conversion for electronic processing and/or storing at the local node. In doing so, the dropped wavelengths become available. The local node may use each of these freed wavelengths to insert local traffic on the available added wavelengths λ_{add}. Note that the dropped wavelengths λ_{drop} and added wavelengths λ_{add} operate at the same optical frequency but carry different traffic (locally dropped and added traffic, respectively). Subsequently, all M wavelengths are combined onto a common outgoing fiber by using an $M \times 1$ wavelength Multiplexer (MUX). The composite optical WDM comb signal may be amplified by using another optical amplifier at the output fiber (e.g., EDFA).

Reconfigurable Optical Add–Drop Multiplexer

In conventional OADMs, the optical add, drop, and in-transit paths are fixed. Hence, conventional OADMs are static and are able to add and drop only a pre-specified set of wavelengths, without providing any possibility to (re)-configure this set. Figure 34.2 depicts a ROADM based on optical cross-bar switches with a single fiber link carrying two wavelengths. ROADMs become reconfigurable by using optical cross-bar switches on the in-transit paths between wavelength DEMUX and MUX. The cross-bar switch has two input ports and two output ports and is in either of two states at any given time: (1) *bar* or (2) *cross* state. In the bar state, the optical in-transit signal is forwarded to the opposite output port without being dropped locally. In this case, the local node is unable to add traffic. In the cross state, the optical in-transit signal is locally dropped. This allows the local node to add its own traffic, which is forwarded to the wavelength MUX and finally onto the common outgoing fiber link. Thus, optical cross-bar switches enable either optical bypassing or add/drop operations. As shown in the figure, the state of each cross-bar switch is typically controlled electronically. In general, the cross-bar switches are controlled independently from each other such that each wavelength of the arriving WDM comb can be accessed separately by the local node.

Optical Cross-Connect

The generic structure of an OXC with N input/output fiber links, each carrying M different wavelength channels, is shown in Figure 34.3. An OXC is an $N \times N \times M$ component with N input fibers, N output fibers, and M wavelength

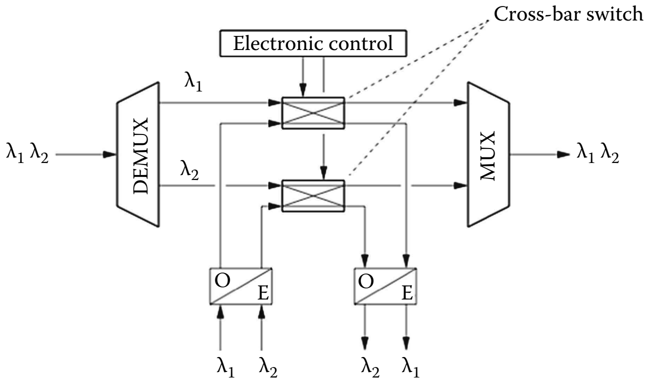

FIG. 34.2
Reconfigurable optical add–drop multiplexer (ROADM).

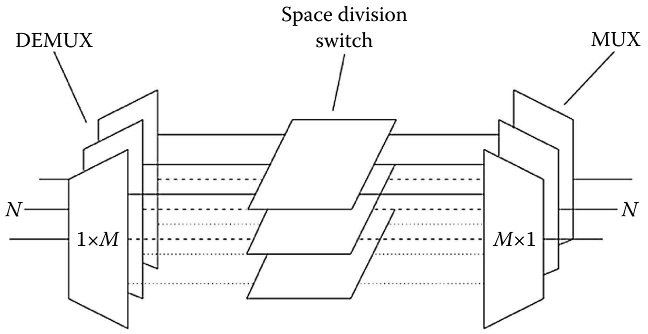

FIG. 34.3
Optical cross-connect (OXC).

channels on each fiber. A DEMUX is attached to each input fiber (and optionally also an optical amplifier, similar to the previously discussed OADM). Each output from a DEMUX goes into a separate wavelength layer. Each wavelength layer has a space division switch that directs each wavelength channel to a selected MUX. Each MUX collects light from M space division switches and multiplexes the wavelengths onto an output fiber. OXCs improve the flexibility and survivability of networks. They provide restoration and can reconfigure the network to accommodate traffic load changes and to compensate for link and/or node failures. An AON that deploys OXCs and ROADMs is commonly referred to as an *optical transport network* (OTN). OTNs are able to provide substantial cost savings due to their flexibility, optical bypass capability, reconfigurability, and restoration [5].

WAVEBAND SWITCHING

Compared to ordinary OXCs, the so-called *multigranularity optical cross-connects* (*MG-OXCs*) support multiple switching granularities apart from wavelength switching. MG-OXCs hold great promise to reduce the complexity and costs of OXCs significantly by switching fibers and wavebands as an entity without demultiplexing the arriving WDM comb signal into its individual wavelengths, giving rise to *waveband switching* (*WBS*). WBS has been receiving considerable attention for its practical importance in reducing the size and complexity of OXCs. Due to the rapid development and worldwide deployment of dense wavelength division multiplexing (DWDM) technologies, current fibers are able to carry hundreds of wavelengths. Using ordinary wavelength-switching cross-connects would require a large number of ports. WBS comes into play here with the promise to reduce the port count, control complexity, and reduce the cost of OXCs. The rationale behind WBS is to group several wavelengths together as a waveband and switch the waveband optically using a single input and a single output port instead of multiple input/output ports, one for each of the individual wavelengths of the waveband. As a result, the size of ordinary OXCs that traditionally switch at the wavelength granularity can be reduced, including the associated control complexity and cost [6].

Multigranularity Optical Cross-Connect

Figure 34.4 depicts a typical multigranularity optical cross-connect consisting of a three-layer MG-OXC and a *digital cross-connect* (DXC) [6]. An MG-OXC not only switches traffic at multiple levels of granularities such as fiber, waveband, and wavelength levels but also allows to add and to drop traffic at multiple granularities by deploying a bank of transmitters and receivers. Traffic can be shifted from one granularity level to another by using appropriate MUXs and DEMUXs, as described in greater detail shortly. Note that for subwavelength switching (e.g., time division multiplexing [TDM] switching and grooming), the MG-OXC is equipped with an additional DXC that performs the OEO conversion. More specifically, the MG-OXC of Figure 34.4 consists of the fiber cross-connect (FXC), band cross-connect (BXC), and wavelength cross-connect (WXC) layers. All three layers use add and drop ports to locally add and drop traffic at the corresponding granularity level, respectively. Unlike the FXC layer, both the BXC and WXC layers deploy MUXs and DEMUXs. The WXC layer comprises a WXC switch fabric that is used to switch wavelengths, that is, light paths, and one or more W_{add}/W_{drop} ports to add/drop wavelengths locally. Furthermore, the WXC layer makes use of one or more band-to-wavelength (BTW) DEMUXs to demultiplex wavebands into wavelengths and one or more wavelength-to-band (WTB) MUXs to multiplex wavelengths into wavebands. Similarly, the BXC layer comprises a BXC switch fabric for waveband switching and one or more B_{add}/B_{drop} ports to add/drop wavebands locally. In addition, the BXC layer comprises one or more fiber-to-band (FTB) DEMUXs to demultiplex fibers into wavebands and one or more band-to-fiber (BTF) MUXs to multiplex wavebands into fibers. The various MUXs and DEMUXs are used for transporting traffic from one layer to another. Finally, the FXC layer consists of an FXC switch fabric for space switching of fibers.

In contrast to an ordinary single-granularity OXC that switches each wavelength individually using a separate port, the MG-OXC is able to switch a fiber using one port if none of the corresponding wavebands and wavelengths needs to be dropped or added. Otherwise, the fiber will be demultiplexed into wavebands. Only those wavebands whose wavelengths need to be added or dropped are further demultiplexed. Each of the remaining wavebands will be switched as an entire waveband using one port at the BXC layer. In doing so, fibers and wavebands that carry bypass traffic do no need to undergo demultiplexing and multiplexing and can be switched as an entity, resulting in a reduced number of required ports compared to ordinary wavelength-switching OXCs. Thus, relatively small-scale modular switch fabrics are sufficient to build scalable MG-OXCs.

At the downside, MG-OXCs suffer from a few drawbacks. First, besides the switch fabrics, each MG-OXC requires

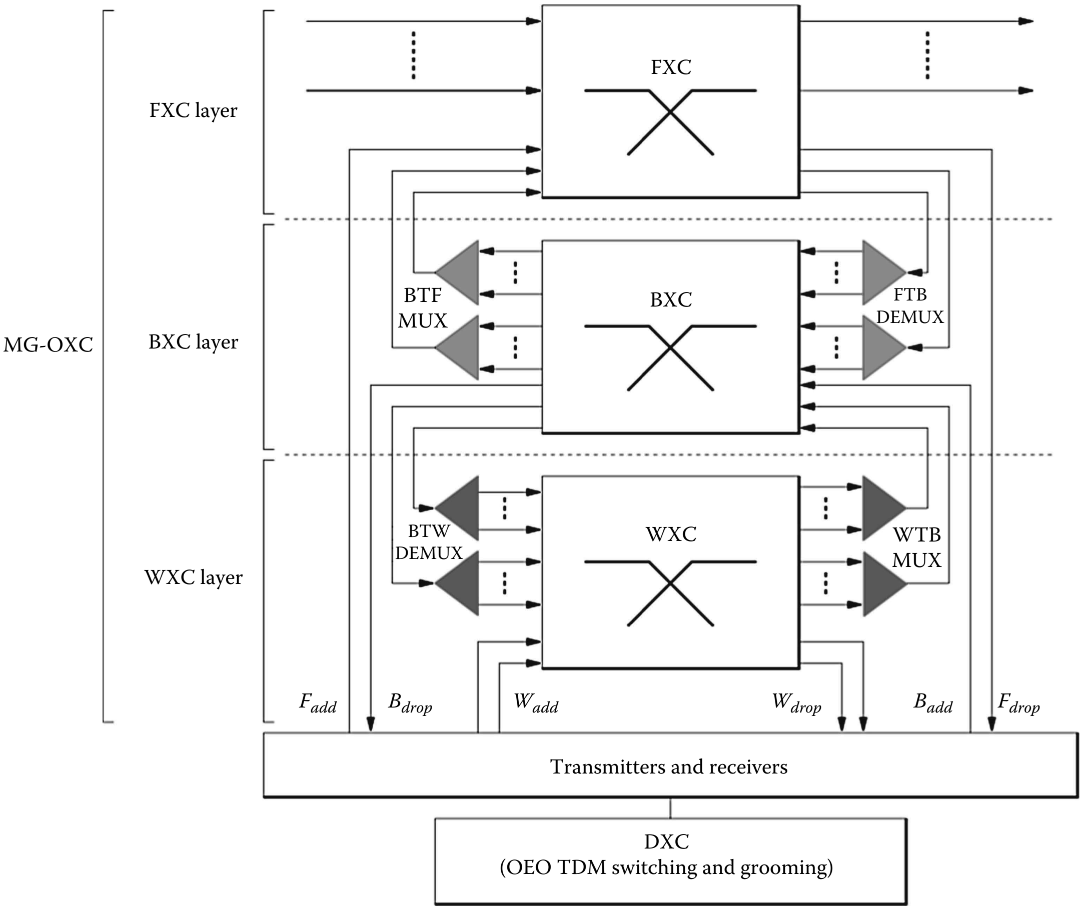

FIG. 34.4
Multigranularity optical cross-connect (MG-OXC). (After Cao, X., Anand, V., and Qiao, C., Waveband switching in optical networks, IEEE Communications Magazine, 41(4), 105, April 2003 © (2003) IEEE.)

additional DEMUXs to shift traffic between the different layers. Second, the optical signal quality may deteriorate significantly within each MG-OXC due to the fact that lightpaths may need to go through multiple layers. Both shortcomings can be mitigated by using single-layer MG-OXCs instead of multilayer MG-OXCs. In single-layer MG-OXCs, all lightpaths traverse only a single switch fabric. Besides complexity, cost, and signal quality issues, it is important to consider the given traffic loads in making a choice between single-layer and multilayer MG-OXCs. It was shown that for static traffic, single-layer MG-OXCs provide a greater reduction in size than multilayer MG-OXCs and vice versa for dynamic traffic [7].

Waveband Grouping

With WBS, it is important to find out how many and which wavelengths need to be grouped together into a single waveband in order to meet certain performance requirements. Toward this end, waveband grouping strategies for WBS networks that satisfy different performance metrics were developed. The existing waveband grouping strategies can be classified into the following two categories: (1) *end-to-end* waveband grouping and (2) *intermediate* waveband grouping. In general, intermediate waveband grouping strategies outperform end-to-end grouping strategies in terms of cost savings. To see this, consider the following illustrative example. Let us assume that there are three geographically distributed source nodes, each using a separate wavelength to send traffic to the same destination node *D*. Through intermediate waveband grouping, the three wavelengths can be grouped into a single waveband at any common intermediate node *X*, resulting in a reduced number of required ports at MG-OXCs between intermediate node *X* and destination node *D*. On the contrary, end-to-end waveband grouping strategies are unable to group the three wavelengths into a single waveband because the sources are not collocated [8].

TDM Switching and Grooming

As shown in Figure 34.4, MG-OXCs can be equipped with an additional DXC in order to perform TDM switching and grooming in the electrical domain by means of OEO conversion of wavelengths and wavebands. A hybrid optoelectrical switch architecture that integrates all-optical fiber and waveband switching and electrical TDM switching was investigated in Ref. [9]. The proposed switch architecture does not support wavelength switching, but it combines the scalability and cost savings of WBS with the flexibility of subwavelength TDM switching in the electrical domain. In addition, the electrical TDM switch can be used to perform wavelength conversion and multicasting. A similar hybrid optoelectrical switch architecture that integrates all-optical waveband switching and electrical traffic grooming was investigated in Ref. [10]. The presented design study of WBS networks based on such hybrid WBS-OEO grooming switch architectures also takes physical transmission impairments on all-optical waveband switching paths into account (e.g., fiber nonlinearities, amplified spontaneous emission [ASE], and polarization mode dispersion [PMD]). Those impairments limit the maximum distance and number of nodes the optical signal can traverse without undergoing OEO conversion.

PHOTONIC SLOT ROUTING

To avoid the loss of optical transparency and the need for electronic traffic grooming, *photonic slot routing* (PSR) was proposed to allow entire slots, each carrying multiple packets on distinct wavelength channels, to be switched all-optically and individually [11]. Unlike lightpath-based optical networks, PSR allows traffic aggregation to be done optically by intermediate nodes (i.e., without electronic traffic grooming) in order to use the bandwidth of each wavelength channel efficiently.

Photonic Slot

In PSR networks, time is divided into fixed-size slots. Each slot spans all wavelengths in the PSR network with both slot boundaries aligned across all wavelengths. Correspondingly, the resultant multiwavelength slot is called *photonic slot*. Each wavelength in the photonic slot contains a single data packet. Thus, given W wavelength channels, the photonic slot may carry up to W data packets. Furthermore, all packets in a given photonic slot are required to be destined for the same node, but each photonic slot may be destined for a different node. Due to the fact that all packets must have the same destination, the photonic slot can be routed as a single entity by intermediate nodes, thereby avoiding the need for demultiplexing the individual wavelengths and routing them individually. By requiring that photonic slots are routed as a single entity, PSR networks exhibit a number of advantages. First, no wavelength-sensitive components are required at intermediate nodes. Instead, wavelength-insensitive components can be used, resulting in lower network costs and avoidance of interchannel switching crosstalk in adjacent wavelength channels since wavelengths are not separated at intermediate nodes. Second, PSR reduces the complexity of the switching operation and its electronic control at intermediate nodes by a factor of W. Third, PSR nodes can be built in a cost-effective fashion by using relatively simple optical components based on proven technologies. At the downside, the merits of PSR come at the expense of achieving and maintaining slot synchronization network wide.

As shown in Figure 34.5, PSR network nodes may perform the following three functions on a per-photonic-slot basis [12]:

Photonic slot switching: Photonic slots arriving on any input port of a given PSR node can be switched individually to any output port. Photonic slots that are switched to the same output port simultaneously cause contention at the output port. Several approaches exist to resolve contention in PSR networks, as discussed in greater detail shortly.

Photonic slot copying: A photonic slot arriving on an input port of a given PSR node is duplicated and switched to two or more output ports. Photonic slot copying can be used to realize multicasting of photonic slots (i.e., transmitting a photonic slot to multiple destination nodes). Note that this function may be exploited by a source node to transmit packets in the same photonic slot even though the packets are not intended for the same destination.

Photonic slot merging: Photonic slots concurrently arriving on multiple input ports of a given PSR node are switched to the same output port, thus overlapping with one another to form a new single photonic slot that departs from the PSR node. Clearly, the photonic slot merging function is only permissible if the merging slots are compatible in the sense

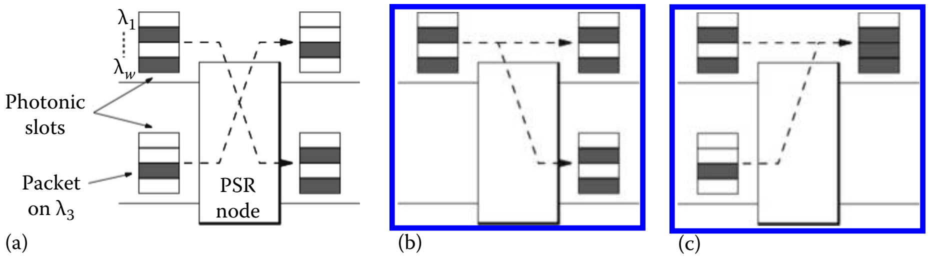

FIG. 34.5

Photonic slot routing (PSR) functions: (a) photonic slot switching, (b) photonic slot copying, and (c) photonic slot merging. (After Elek, V., Fumagalli, A., and Wedzinga, G., Photonic slot routing: A cost-effective approach to designing all-optical access and metro networks, IEEE Communications Magazine, 39(11), 164, November 2001 © (2001) IEEE.)

that they do not carry packets on the same wavelengths. This function allows packets coming from different input ports to be forwarded together toward a common destination.

Contention Resolution

One way to completely avoid contention at PSR nodes network wide is the use of TDM. Recall that in PSR networks, all nodes need to be synchronized. Based on the available network-wide synchronization, TDM can be used to allocate bandwidth to the contention-free transmission of photonic slots. With TDM, transmission on each link takes place in predefined, periodically recurring time frames, which comprise an equal number of time slots. In each time frame, connections each with a transmission capacity of one packet per frame can be established between any pair of source and destination PSR nodes [11].

Figure 34.6 shows the architecture of a PSR node that can be used as a source or intermediate PSR node in mesh WDM networks [13]. The PSR node consists of a wavelength-insensitive optical packet switch, switched delay lines (SDLs) to temporarily store photonic slots at intermediate PSR nodes, and electronic buffers to hold locally generated packets at source PSR nodes, one buffer for each destination. An additional control wavelength is used in each photonic slot that carries the header of the photonic slot. The header contains information such as the destination of the photonic slot and which wavelengths in the photonic slot are occupied by packets. At each input fiber, a DEMUX is used to extract the header of each photonic slot. Since it takes some time to process the header and to configure the optical packet switch, fiber delay lines (FDLs) are used to delay the arriving photonic slots. On a given output fiber, the PSR node may insert packets into photonic slots, which are destined for the appropriate destination by using couplers. For instance, in Figure 34.6 a departing photonic slot carries packets on wavelengths λ_1 and λ_2. The PSR node may insert a packet into the photonic slot on wavelength λ_3.

Several schemes to resolve contention were examined in Ref. [13]. Apart from retransmitting dropped photonic slots by source PSR nodes and buffering contending photonic slots in SDLs at intermediate PSR nodes, the benefit of *deflection routing* was investigated to resolve contention in mesh PSR networks. With deflection routing, in the event of contention a photonic slot is routed to some nonbusy link other than the link specified by the routing algorithm in use (e.g., shortest path first [SPF]). Thus, if two photonic slots are contending, one slot is routed through the output fiber link specified by the routing algorithm whereas the other slot is routed through any of the remaining free output fiber links. To prevent photonic slots from deflecting too often and thereby consuming too many network resources, photonic slots will be dropped (and need to be retransmitted) if they have been deflected up to a predefined number of times. Toward this end, a deflection counter is placed in the header of each photonic slot. Each time a photonic slot is deflected, the counter is decremented. For

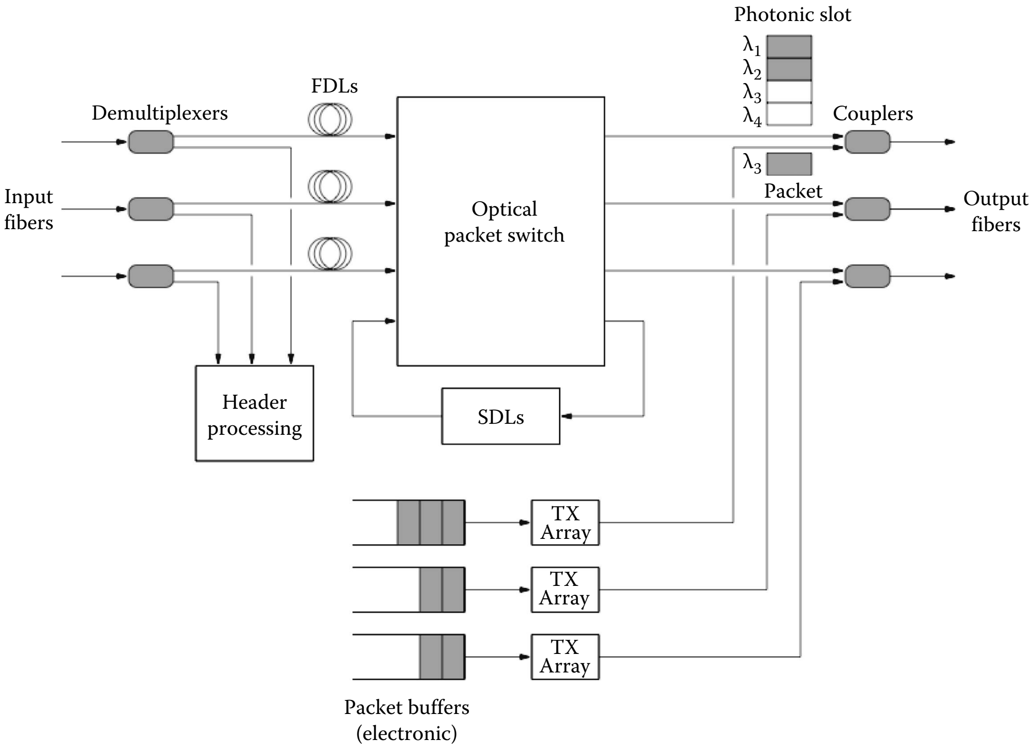

FIG. 34.6
PSR node architecture. (After Zang, H., Jue, J.P., and Mukherjee, B., Capacity allocation and contention resolution in a photonic slot routing all-optical WDM mesh network, IEEE/OSA Journal of Lightwave Technology, 18(12), 1728, December 2000 © (2000) IEEE.)

shortest routing, the performance of the two contention resolution techniques, buffering and deflection routing, is similar under light traffic loads. This is because at low load, there is only little contention and therefore no buffering or deflection routing is necessary. With increasing load, deflection routing may outperform buffering. This is due to the fact that shortest path routing leads to unbalanced link loads. Deflection routing is able to balance the load on the links by using alternate paths to each destination beside the shortest path.

It is important to note that the benefits of deflection routing heavily depend on the network topology and the applied routing algorithm. It might occur that deflection routing actually performs worse than buffering due to the excess consumption of network resources by deflected photonic slots. Generally, deflection routing becomes beneficial if there are one or more alternate paths to the destination that are less loaded than the path specified by the selected routing algorithm. By applying adaptive (dynamic) routing algorithms instead of fixed shortest path routing, the load can be better balanced across network links, resulting in reduced contention at intermediate PSR nodes and an improved network performance.

OPTICAL FLOW SWITCHING

One of the main bottlenecks in today's Internet is (electronic) routing at the Internet Protocol (IP) layer. Several methods have been proposed to alleviate the routing bottleneck by switching long-duration flows at lower layers. In doing so, routers are offloaded and the electro-optical bottleneck is alleviated. This concept of lower-layer switching can be extended to switching large transactions and/or long-duration flows at the optical layer, giving rise to *optical flow switching* (OFS) [14]. A flow denotes a unidirectional sequence of IP packets between a given pair of source and destination IP routers. Both source and destination IP addresses, possibly together with additional IP header information such as port numbers and/or type of service (ToS), are used to identify a flow. In OFS, a lightpath is established for the transfer of large data files or long-duration and high-bandwidth streams across the network. The simplest form of OFS is to use an entire wavelength for a single transaction. Alternatively, flows with similar characteristics may be aggregated and switched together by means of grooming techniques in order to improve the utilization of the setup lightpath. OFS helps reduce the computation load and processing delay at IP routers, but it also raises several issues that need to be addressed. For instance, how can the start and end of flows be recognized? Also, since in OFS the set up of a lightpath takes at least one round-trip time between source and destination IP routers, the size of a flow should be in the order of the product of round-trip propagation delay and line rate at which the lightpath operates.

Figure 34.7 illustrates the difference between OFS and conventional electronic routing by IP routers [15]. In OFS, a

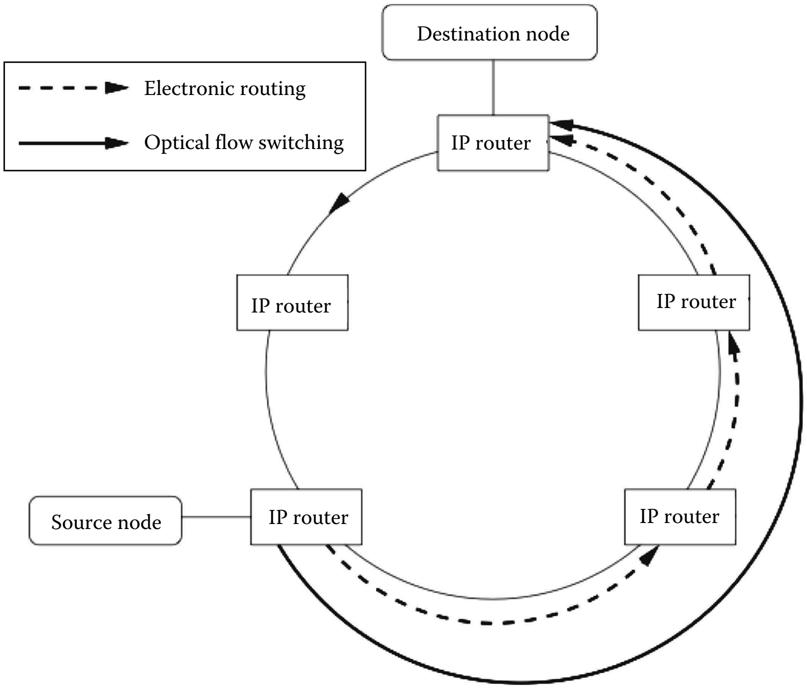

FIG. 34.7

Optical flow switching (OFS). (After Froberg, N.M., Henion, S.R., Rao, H.G. et al., The NGI ONRAMP test bed: Reconfigurable WDM technology for next generation regional access networks, IEEE/OSA Journal of Lightwave Technology, 18(12), 1697, December 2000 © 2000 IEEE.)

large set of data is routed all-optically in order to bypass and thereby offload intermediate IP routers in the data path and their associated queues. Toward this end, a dedicated lightpath is established from the source node to the destination node, allowing the source node to transmit its large data set directly to the destination node. The set up lightpath eliminates the need for packet processing at intermediate routers (e.g., buffering and routing). It is important to note that OFS can be end-user initiated or IP-router initiated. In the first case, the end user has a large amount of data to send and requests a lightpath to the corresponding destination node. In the second case, an IP router detects large volume flows and diverts them to a lightpath. OFS is practical if the flow lasts at least as long as it takes to set up the lightpath. With current technology and taking typical propagation delays into account, the duration of a flow must be on the order of a few up to tens of milliseconds in order to justify the round-trip time to set up the lightpath.

OFS provides several advantages. Apart from mitigating the aforementioned electro-optical bottleneck by optically bypassing and thus offloading electronic IP routers, OFS represents the highest-grade quality of service (QoS) since the established lightpath provides a dedicated connection that is not impaired by the presence of other users. At the downside, OFS must carefully determine when to set up a lightpath since wavelengths are typically a scarce resource that cannot be arbitrarily assigned. Without the use of wavelength converters, it is necessary to assign the same wavelength to a lightpath along its entire path. This wavelength continuity constraint further restricts the number of available wavelengths.

To enable dynamic lightpath setup in OFS, several important issues must be addressed. These issues include flow routing, wavelength assignment, and connection set up. Routing involves selecting an appropriate path; wavelength assignment involves assigning an appropriate wavelength on each link of the selected path; and connection set up involves signaling to create the lightpath, given a route and a wavelength on each link along the route. Under the assumption that a reliable control network for signaling is available and flow requests are generated by end users, the following two integrated OFS approaches exist for flow routing: wavelength assignment and connection setup [16].

Tell-and-Go Reservation

The so-called *tell-and-go* (*TG*) approach is a distributed algorithm with no wavelength conversion. TG is based on link state updates. Received updates are processed at each network node to acquire and maintain the global network state. In TG, the global network state may contain inaccurate information due to the latency involved in sending updates. Given the network state, TG uses a combined routing and wavelength assignment strategy. For any routing decision, the *K* shortest paths are considered. Routing proceeds by using first-fit wavelength assignment over the *K* paths. First-fit wavelength assignment numbers all wavelengths in sequential order and routes an optical flow over the path with the lowest-numbered wavelength available at each hop. If no route with an available wavelength can be found, the optical flow is dropped. Connection setup is achieved using the TG principle, whereby a control packet precedes the optical flow along the chosen route. The control packet establishes a lightpath for the trailing optical flow. Note that TG represents a one-way reservation, where the control packet travels only in the direction from source to destination and not backward again to the source, as is done in a two-way reservation. If at any intermediate node sufficient resources are not available, the control packet and the corresponding optical flow are dropped.

Reverse Reservation

In the so-called *reverse reservation* (*RR*) approach, the initiator of an optical flow sends information-gathering packets, termed info-packets, to the destination node on the *K* shortest paths. These info-packets record the link state information at each hop. Upon arrival at the destination, the info-packets contain information about the link state of all links along the *K* routes. Routing and wavelength assignment are performed by the destination node, once all *K* info-packets from the sender have arrived. The calculation is done using the first-fit wavelength assignment, as described earlier. Similar to TG, if no route with an available wavelength can be found, the optical flow is dropped. Once a route has been selected by the destination, a reservation control packet is sent along the chosen route in reverse to establish the connection. This reservation control packet configures intermediate switches along the selected route for the chosen wavelength and finally informs the initiator that the lightpath has been set up successfully for sending the optical flow. Otherwise, if the reservation control packet does not find sufficient available resources along the selected route, the reservation is dropped and all resources held by the reservation in progress are released by sending additional control packets. Note that unlike TG, RR does not require (periodic or event-driven) updates to acquire and maintain global network state.

OPTICAL BURST SWITCHING

An *optical burst switching* (OBS) network consists of OBS nodes interconnected by WDM fiber links, as shown in Figure 34.8. At the edge of the OBS network, one or more users are attached to an OBS node. Typically, a user at the edge of the OBS network consists of an electronic IP router equipped with an OBS interface, while an OBS node inside the OBS network comprises an optical switching fabric, a switch control unit, and routing and signaling processors. Following the approach in Ref. [17], the OBS network and its operation are best explained by discussing first the functions executed by users at the edge of the OBS network, followed by a description of the functions carried out by OBS nodes inside the OBS network.

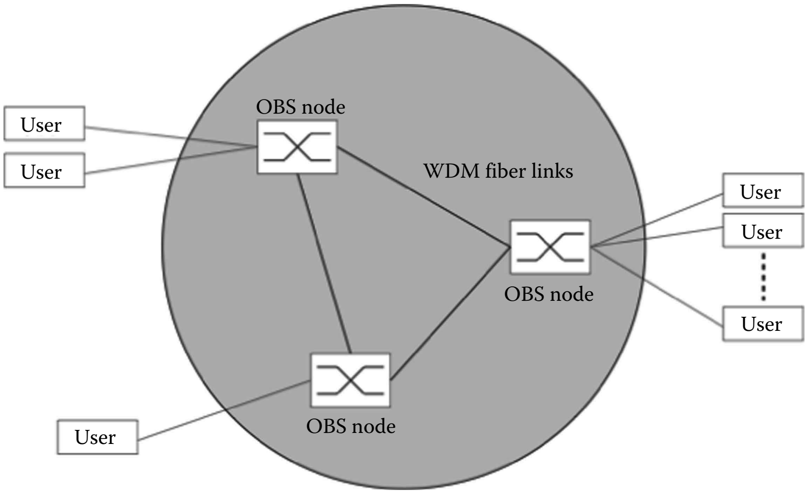

FIG. 34.8

Optical burst switching (OBS) network architecture. (After Battestilli, T. and Perros, H., An introduction to optical burst switching, IEEE Communications Magazine, 41(8), S10, August 2003 © 2003 IEEE.)

OBS Network Edge

The four functions executed by edge OBS users can be categorized into: (1) *burst assembly*, (2) *signaling*, (3) *routing and wavelength assignment*, and (4) the computation of the *offset* time for the control packet, which is sent prior to the data burst. These four functions are explained in detail next.

Burst Assembly

OBS users collect traffic originating from upper network layers (e.g., IP layer), sort it based on destination addresses, and aggregate it into variable-size data bursts by using appropriate burst assembly algorithms. The choice of the burst assembly algorithm can have a significant impact on the performance of the OBS network. Several burst assembly algorithms have been proposed and investigated up to date, as discussed in greater detail shortly. Generally, burst assembly algorithms have to take the following parameters into account: timer, minimum burst size, and maximum burst size. The timer is used by the OBS user to determine when to assemble a new burst. The minimum and maximum burst size parameters determine the length of the assembled burst. These parameters must be carefully set since long bursts hold network resources for long time periods and may therefore cause higher burst losses, while short bursts cause an increased number of control packets since each burst is preceded by a separate control packet. Padding may be used by the burst assembly algorithm if an OBS user does not have enough traffic to assemble a minimum-size burst when the timer expires. It is interesting to note that burst assembly helps reduce the degree of self-similarity of higher layer traffic. Reducing self-similarity is a desirable property since it makes the traffic less bursty and thus leads to a decreased queuing delay and smaller packet loss in OBS networks.

Signaling

Signaling is used to set up a connection in the OBS network for an assembled burst between a given pair of source and destination edge OBS users. In OBS networks, there are two types of signaling: (1) *distributed signaling with one-way reservation* and (2) *centralized signaling with end-to-end reservation*.

Most of the proposed OBS network architectures use the distributed signaling with one-way reservation, whose time–space diagram is shown in Figure 34.9. In the one-way reservation scheme, a source OBS user sends a control packet on a separate out-of-band control channel to the ingress OBS node, say, OBS node *A*, prior to transmitting the corresponding burst. The out-of-band control channel may be a dedicated signaling wavelength channel or a separate control network. The control packet contains information about the burst (e.g., size). The control packet is OEO converted and processed in the electronic domain at each intermediate OBS node along the path to the destination OBS user. After a prespecified delay, called *offset*, the corresponding data burst is sent on one of the available data wavelength channels without waiting for an acknowledgment (ACK) that the connection between source and destination OBS users has been established successfully. Compared to the conventional two-way reservation scheme with ACK, the resultant one-way reservation significantly decreases the connection setup time, which effectively equals the one-way, end-to-end propagation delay plus the time required to electronically process the control packet and configure the optical switch fabric at all intermediate OBS nodes. In the one-way reservation scheme, however, control packets may not be successful in setting up connections due to congestion on the data wavelength channels. As a result, the optical switch cannot be configured as needed and the corresponding burst will be dropped. Thus, in OBS networks with one-way reservation, there is a certain nonzero burst loss probability. Typically, the retransmission

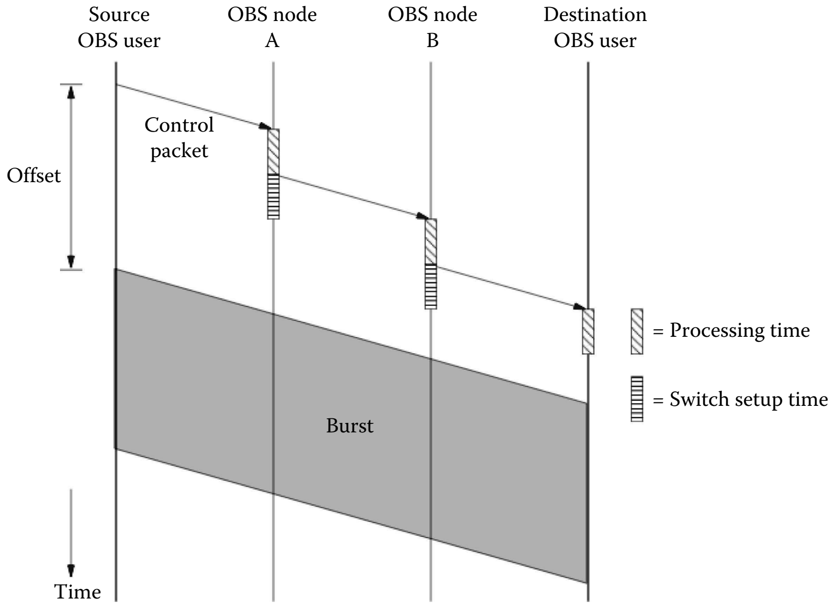

FIG. 34.9
Distributed OBS signaling with one-way reservation. (After Battestilli, T. and Perros, H., An introduction to optical burst switching, IEEE Communications Magazine, 41(8), S10, August 2003 © 2003 IEEE.)

of lost bursts is not provided in OBS networks but is left to higher layer protocols.

In the second less frequently used OBS signaling approach, namely, centralized signaling with end-to-end reservation, OBS users send their connection setup requests to ingress OBS nodes, which in turn inform a central request server about the received connection setup requests. The central server has global knowledge about the current status of all OBS nodes and data wavelength channels throughout the OBS network. Based on this global knowledge, the central server processes all received connection setup requests and sends ACKs to the requesting edge OBS users. Upon the receipt of their ACKs, OBS users transmit their bursts.

Routing and Wavelength Assignment

Routing in OBS networks can be done either on a hop-by-hop basis using fast routing table lookup algorithms at intermediate OBS nodes (similarly to IP networks) or by deploying *generalized multiprotocol label switching* (GMPLS) routing protocols to compute explicit or constraint-based routes at edge OBS users. Along the selected path, each link must be assigned a wavelength on which bursts are carried. In OBS networks, wavelength assignment with and without wavelength conversion at intermediate OBS nodes is possible, whereby wavelength conversion may be fixed, limited range, full range, or sparse.

Offset

As depicted in Figure 34.9, one of the salient features of OBS is the fact that an edge OBS user sends a control packet on a separate control channel prior to sending the corresponding data burst on one of the available data wavelength channels. More precisely, after sending the control packet, an OBS user waits for a fixed or variable offset time until it starts transmitting the corresponding burst. The offset is used to let the control packet be processed, reserve the required resources, and configure the optical switching fabric at each intermediate OBS node along the selected path before the corresponding burst arrives. If the control packet is successful in reserving the needed resources and configuring the switch, the arriving burst can *cut through* each intermediate OBS node without requiring any buffering or processing. Clearly, the estimation and setting of the offset time is crucial. Ideally, the offset estimation should be based on the number of traversed OBS nodes and the processing and switch setup times of each of them. In practice, however, the number of intermediate OBS nodes may not be known to the source OBS user or may change over time. Furthermore, the current level of congestion in the OBS network needs also to be taken into account in the offset estimation in order to achieve an acceptable burst loss probability. If the offset is estimated incorrectly, the burst may arrive at an OBS node before the optical switch has been configured properly. Clearly, the offset estimation in OBS networks is of utmost importance to achieve high resource utilization and low burst loss.

OBS NETWORK CORE

OBS nodes located in the core of OBS networks perform the following two functions: (1) *scheduling* of resources and (2) *contention resolution*, if there are not enough resources to successfully schedule simultaneously arriving bursts. Both scheduling and contention resolution are discussed below.

Scheduling

Each control packet contains information about the corresponding burst (e.g., offset and size). Based on this information, OBS nodes schedule the resources inside the local optical switching fabrics such that bursts can cut through them. Based on the start and end times a burst occupies the resources inside the optical switching fabric of an OBS node, the resource scheduling schemes proposed for OBS networks can be classified as follows:

- *Explicit setup:* In explicit setup, a wavelength channel is reserved and the optical switching fabric is configured immediately after receiving and processing the control packet.
- *Estimated setup:* In estimated setup, the reservation of a wavelength channel is delayed and, by using information provided by the control packet, the OBS node estimates the arrival time of the corresponding burst. A wavelength channel is reserved and the optical switching fabric is configured at the OBS node right before the estimated arrival time of the burst.
- *Explicit release:* In explicit release, the source OBS user sends an additional trailing control packet to indicate the end of the burst. After receiving the trailing control packet, the OBS node releases the reserved wavelength.
- *Estimated release:* In estimated release, an OBS node estimates the end of the burst by using the offset and size information carried in the preceding control packet. Based on this information, an OBS node is able to estimate the time when to release the reserved wavelength after the corresponding burst has passed through the switching fabric.

There exist four possibilities to combine these four OBS resource scheduling schemes: explicit setup/explicit release, explicit setup/estimated release, estimated setup/explicit release, and estimated setup/estimated release. Each combination provides a different performance-complexity tradeoff. Clearly, the estimated setup/release schemes provide improved resource utilization over their explicit counterparts. However, the explicit setup/release schemes are simpler to implement, but they occupy the wavelength channels for longer time periods and therefore may lead to an increased burst loss probability.

Note that the choice of the resource scheduling scheme in OBS networks also depends on the burst assembly algorithm used by the OBS users at the OBS network edge. To see this, consider the following example. If an edge OBS user first assembles a burst and then sends a control packet with information about the offset and size of the burst, OBS nodes are able to apply estimated setup and estimated release schemes. However, if the control packet is sent before the corresponding burst is completely assembled, OBS nodes have to deploy explicit release schemes (i.e., the edge OBS user has to send another trailing control packet, which indicates the end of the burst transmission).

Contention Resolution

Contention resolution is one the main design objectives in OBS networks. Contention occurs if a burst arrives at an OBS node and all local resources are occupied or if two or more simultaneously arriving bursts contend for the same resource. Several techniques can be employed by OBS nodes to resolve contention. Contention resolution techniques may be applied in the time, wavelength, or space domains or any combination thereof. Among others, contention among simultaneously arriving bursts may be resolved by using FDLs where bursts are temporarily stored such that a given output port of an OBS node contending bursts can be sent out sequentially without collision. Another way to resolve contention is deflection routing, where one burst is forwarded through the intended output port of an OBS node while the remaining contending bursts are sent through other output ports along different paths. Wavelength conversion is another approach to resolve contention where bursts leave the same output port of an OBS node on different wavelength channels at the same time.

OBS Medium Access Control Layer

OBS introduces a number of specific network design issues that must be addressed properly in order to implement the aforementioned functions. To implement these functions, a medium access control (MAC) layer is required between the IP layer and the optical layer [18]. Figure 34.10 depicts the functional blocks needed at the OBS MAC layer as well as at the optical layer for implementing OBS networks. The functional blocks correspond to the aforementioned functions that need to be executed by edge OBS users and core OBS nodes at the MAC layer and optical layer, respectively. At the OBS MAC layer, source and destination OBS users located at the edge of the OBS network perform the functions of burst assembly/disassembly, offset computation, control packet generation, routing and wavelength assignment, and signaling. At the optical layer, intermediate core OBS nodes are responsible for scheduling and contention resolution of in-transit bursts. The OBS MAC layer together with the underlying optical layer offers services that guarantee certain burst blocking probabilities to the higher-layer client IP routers.

OPTICAL PACKET SWITCHING

Most of the concepts used in *optical packet switching* (*OPS*) networks are borrowed from their electronic packet switching counterparts such as asynchronous transfer mode (ATM)

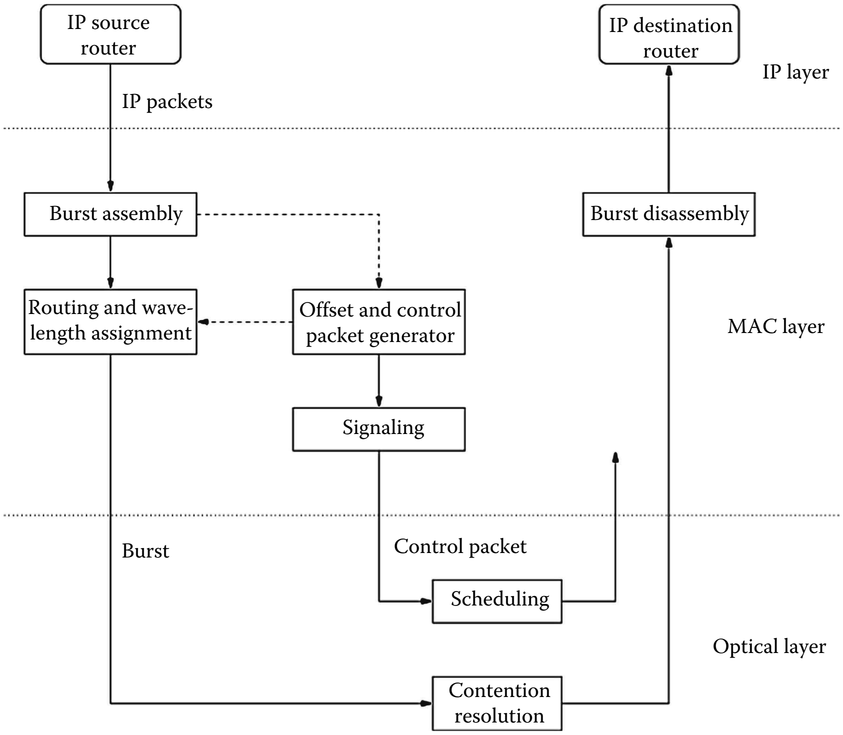

FIG. 34.10
Block diagram of OBS networks consisting of IP, MAC, and optical layers. (After Verma, S., Chaskar, H., and Ravikanth, R., Optical burst switching: A viable solution for terabit IP backbone, IEEE Network, 14(6), 48, November/December 2000 © 2000 IEEE.)

switches and IP routers. OPS may be viewed as an attempt to mimic electronic packet switching in the optical domain while taking the shortcomings and limitations of current optics and photonics technology into account. The challenge faced in OPS involves developing an elegant solution to the mismatch between the transmission capacity offered by the WDM optical layer and the processing power of electronic switches and routers [19]. This mismatch, also referred to as the electro-optical bottleneck, can be in principle alleviated by increasing the degree of parallelism at the electronic layer. In practice, however, a higher level of parallelism not only increases the complexity of electronic switches and routers but also results in a significantly increased power consumption, cost, and footprint. By using low-cost optics and photonics technology and performing part of the switching and/or routing in the optical domain, electronic switches and routers can be offloaded, resulting in reduced complexity, footprint, and power consumption, improved performance, as well as significant cost savings.

At present, the implementation of OPS where both the packet header and payload are processed and routed entirely in the optical domain is challenging due to the lack of optical random access memory (RAM) and the difficulty to execute complex computations and logical operations using only optics and photonics without requiring any OEO conversion at OPS routers. An interesting approach that allows practical OPS networks to be realized in the near term is the so-called *optical label switching* (OLS) technique [20]. OLS may be viewed as a particular implementation of OPS. In OLS, only the packet header, referred to as *label*, is processed electronically for routing purposes while the payload remains in the optical domain. The label can be differentiated from the payload in a number of ways (e.g., diversities in time, wavelength, and modulation format). In general, the label is encoded at a lower bit rate compared to the payload in order to simplify OE conversion during the label extraction and rewriting process. The basic functions performed by each OLS router involve the following steps: (1) extraction of the label from the optical packet, (2) electronic processing of the label to obtain routing information, (3) if necessary, optical routing of the payload and resolving contention, and (4) rewriting of the label and recombining it with the optical payload. Several key enabling technologies for OLS networks have been developed over the last few years, including optical label generation, optical label swapping, optical buffering, clock recovery, and wavelength conversion. In the following, we elaborate on how these technologies can be exploited to build future OPS networks.

Packet Format

Figure 34.11 depicts a generic optical packet format consisting of a packet header and payload with additional guard bands before and after the payload [21]. The guard bands are used to cope with timing uncertainties of optical packets arriving at OPS nodes. The payload header should provide

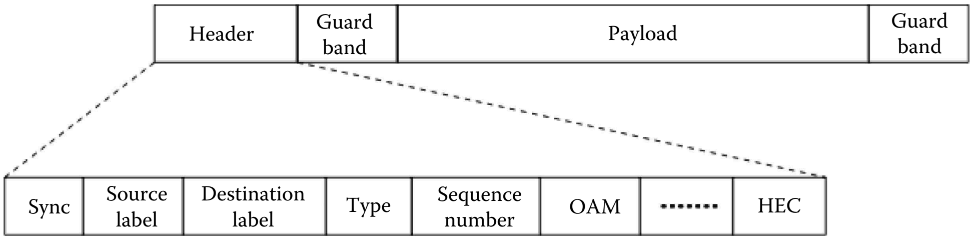

FIG. 34.11
Generic optical packet format. (After El-Bawab, T.S. and Shin, J.-D., Optical packet switching in core networks: Between vision and reality, IEEE Communications Magazine, 40(9), 60, September 2002 © 2002 IEEE.)

a reasonable trade-off between overhead size and the number of supported control functions that allow OPS nodes to route the packet across the OPS network. Among others, the header may comprise the following fields:

- *Sync:* Delineation and synchronization bits
- *Source label:* Source node address
- *Destination label:* Destination node address
- *Type:* Type and priority of packet and carried payload
- *Sequence number:* Packet sequence number to reorder packets arriving out of order and guarantee in-order packet delivery
- *OAM:* Operation, administration, and maintenance functions
- *HEC:* Header error correction

Several methods exist for encoding packet headers, for example, the packet header can be encoded at a lower bit rate than the payload in order to simplify electronic processing of the header. Alternatively, the header can be subcarrier multiplexed (SCM) with the payload. Typically, in practical OPS networks, a small portion of the optical power of an arriving packet is tapped off at each intermediate OPS node. The packet header is OE converted and processed electronically. As mentioned earlier, this kind of implementation of OPS is known as OLS. All-optical packet header processing, which avoids OEO conversion of the header, is still in its infancy and currently allows only for simple operations such as label matching by using *optical correlators*. Optical correlators are used to recognize addresses and work as follows. When an arriving packet's destination address matches the signature of an optical correlator, autocorrelation pulses that are above a given threshold are generated. Otherwise, cross-correlation pulses below the threshold are generated. By deploying simple threshold detection at the output of the optical correlator, the packet's destination address can be recognized and the packet can be forwarded to the appropriate OPS output port.

Switch Architecture

An OPS node will not be restricted to a simple switching matrix. Figure 34.12 depicts a generic functional block diagram of an OPS node. In general, an OPS node has multiple input and output ports and consists of an input interface, switching matrix, buffer, output interface, and an electronic control unit [21]. Each of these blocks performs different functions:

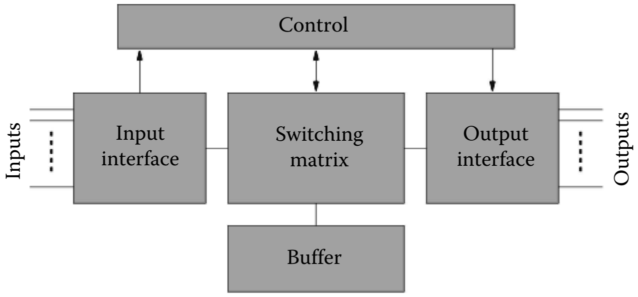

FIG. 34.12
Generic OPS node architecture. (After El-Bawab, T.S. and Shin, J.-D., Optical packet switching in core networks: Between vision and reality, IEEE Communications Magazine, 40(9), 60, September 2002 © 2002 IEEE.)

- *Input interface:* The input interface performs packet delineation to identify the beginning and end of each arriving packet's header and payload. In case of synchronous switches, synchronization is done for performing phase alignment of packets arriving on different wavelength channels and input ports. Note that synchronization is needed only in synchronous switches and not in asynchronous switches. Next, the packet header is extracted, OE converted, decoded, and forwarded to the control unit where it is processed electronically. The control unit processes the routing information in each packet header and configures the switch accordingly. In addition, it updates the header information and forwards the header to the output interface. If necessary, the external wavelength of an arriving optical packet may need to be converted to an internal wavelength for use in the switching matrix.
- *Switching matrix:* The switching matrix performs the switching operation of the payload in the optical domain according to the commands received from the control unit. It also resolves contention by using optical buffers in conjunction with other contention

resolution techniques such as wavelength conversion and deflection routing.

- *Output interface:* The output interface performs 3R (reamplification, reshaping, retiming) regenerative functions to compensate for degradation of the extinction ratio (ER) and signal-to-noise ratio (SNR), power variation between packets, jitter accumulation due to different paths, and insertion losses incurred in the switching matrix. Furthermore, the output interface attaches the updated header to the corresponding optical packet and performs packet delineation. In synchronous switches, the output interface also carries out packet resynchronization. Finally, internal wavelengths are converted back to external wavelengths, if necessary.

Contention Resolution

Similar to OBS networks, a contention may occur in OPS networks whenever two or more packets try to leave an OPS node through the same output port on the same wavelength at the same time. Contention can be resolved by exploiting the time, wavelength, and space dimensions or any combination thereof. More precisely, contention in OPS nodes can be resolved by using buffering (time dimension), wavelength conversion (wavelength dimension), and/or deflection routing (space dimension). Note that deflection routing may be also viewed as a special case of buffering where the OPS network is used as a buffer to store deflection-routed OPS packets. Deflection routing simplifies the OPS node architecture in that no buffers are needed at any OPS node. At the downside, however, deflection-routed packets generally consume more network resources, incur higher delays, and may require packet reordering mechanisms at the destination nodes in order to achieve in-order packet delivery. In the following, we elaborate on the various contention resolution techniques used in OPS networks.

Buffering

According to the position of the optical buffer, OPS nodes can be classified into four major configurations [22]: (1) output buffering, (2) shared buffering, (3) recirculation buffering, and (4) input buffering.

- *Output buffering:* An output-buffered OPS node consists of a space switch with a buffer on each output port, as shown in Figure 34.13a. At any given time, zero or more packets may arrive destined for a particular output port, whereby all of them are placed in the corresponding output buffer. If a packet arrives at a full output buffer, the packet is discarded, resulting in packet loss. Typically, acceptable probabilities for a given packet being lost at a single OPS node are in the range of 10^{-10}–10^{-11}, depending on the application. A packet needs to be output buffered and experiences queuing delay due to contention when more than one packet is destined for the same output port at the same time. Many OPS node architectures are based on output buffering, although OPS nodes usually emulate an output-buffered space switch by means of *virtual output queuing* (VOQ). VOQ is usually deployed in input-buffered OPS nodes, as discussed shortly.
- *Shared buffering:* Shared buffering may be viewed as a form of output buffering, where all output buffers share the same memory space. In doing so, the buffering capacity is not restricted to the number of packets in each individual buffer but to the total number of packets in all buffers together. Shared buffering is commonly used in electronic switches using RAM. In the optical domain, shared buffering may be realized by using FDLs that are shared among all output ports. Shared buffered OPS nodes are able to achieve a significantly reduced packet loss performance with much smaller switch sizes and fewer FDLs than their output-buffered counterparts.
- *Recirculation buffering:* In recirculation buffering, a number of recirculating optical loops from some of the switch output ports are fed back into the switch input ports, as shown in Figure 34.13b. Each optical loop has a certain delay (e.g., one packet). In recirculating buffered OPS nodes, contention is resolved by placing all but one packet into the recirculating loops whenever more than one packet simultaneously arrive

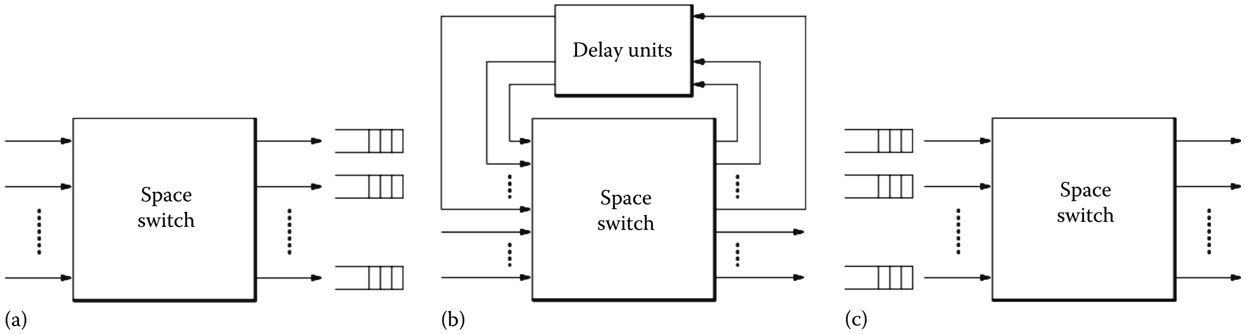

FIG. 34.13
Buffering schemes: (a) output buffering, (b) recirculation buffering, and (c) input buffering.

at the switch input ports destined for the same switch output port. Recirculating packets are forwarded onto the intended switch output port as soon as contention clears. Recirculation buffering helps resolve contention at the expense of optical signal degradation since looped-back optical packets need to pass the delay units and space switch more than once.

- *Input buffering:* Figure 34.13c depicts the final buffering scheme, input buffering. It consists of a space switch with a buffer attached to each separate input port. The fundamental drawback of input buffering is *head-of-line* (HOL) blocking. HOL blocking occurs when the packet at the head of an input queue cannot be forwarded to its intended output port due to current contention. The packet has to be stored in the input queue until there is no more contention at the intended output port. As a consequence, the HOL packet blocks other packets within the same input buffer whose intended output ports are free of contention. Those packets need to wait until the HOL packet is forwarded. As a result, input-buffered OPS nodes suffer from a decreased throughput and increased delay and packet loss. Input buffering is usually never proposed for optical networks, primarily because of its poor performance. In principle, it is possible to enhance input buffering with a so-called *look-ahead capability*, which allows for selecting packets other than those at the head of each input buffer to be forwarded to their corresponding output ports. However, the optical implementation of look-ahead-capable input buffers for OPS networks appear to be too complex. Alternatively, VOQ could be deployed in input-buffered OPS nodes, where each input buffer is replaced with multiple VOQ buffers, one for each switch output port. Arriving optical packets are placed in the corresponding VOQ buffers according to their intended switch output port. Consequently, each VOQ buffer contains only packets destined for the same switch output port, thus avoiding HOL blocking.

Typically, optical buffers are implemented by using an array of FDLs of different lengths or SDLs. Using fibers of variable length to store variable-size optical packets is somewhat tricky since each optical delay line is of fixed length. Once a packet has entered the optical delay line, it can be retrieved only a fixed time period later, equal to the propagation delay of the delay line. This constraint poses some limitations on the realization of optical buffers and resource efficiency of optical variable-size packet switching networks.

All the aforementioned optical buffering schemes can be implemented in either *single-stage* or *multiple-stage* OPS nodes in a *feed-forward* or *feedback* configuration. In a feed-forward configuration, an FDL feeds forward to the next stage of the OPS node. Thus, in a feed-forward configuration, optical packets travel from one end of the OPS node to the other, involving a constant number of FDL traversals. Whereas in a feedback configuration, an FDL sends optical packets back to the input of the same stage, implying that the number of FDL traversals generally differs between optical packets.

Wavelength Conversion

Besides buffering, wavelength conversion is another approach to resolve contention in OPS network nodes by converting optical packets destined for the same output port to different wavelengths. Wavelength conversion can be applied either in conjunction with buffering or without. Figure 34.14 depicts an OPS node architecture that deploys tunable wavelength converters (TWCs) at each input port of the space switch and optical output buffers realized by means of FDLs of different lengths [23]. The OPS node consists of three main blocks.

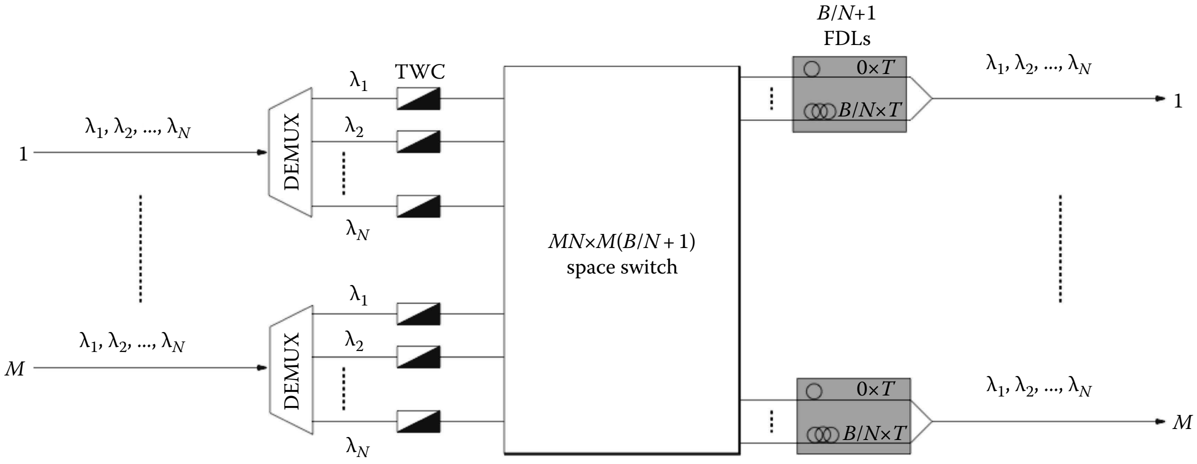

FIG. 34.14

OPS node architecture with input tunable wavelength converters (TWCs) and output FDLs. (After Danielsen, S.L., Hansen, P.B., and Stubkjaer, K.E., Wavelength conversion in optical packet switching, IEEE/OSA Journal of Lightwave Technology, 16(12), 2095, December 1998 © 1998 IEEE.)

The first block uses a DEMUX at each of the M input fibers to separate optical packets arriving on wavelengths $\lambda_1, \lambda_2, \ldots, \lambda_N$, and feeds them into separate TWCs, one for each wavelength; the TWCs address free space available in the optical output buffers by converting incoming packets to the appropriate wavelengths. The second block consists of a nonblocking space switch with $M \cdot N$ input ports and $M(B/N + 1)$ output ports, where B is assumed to be an integer multiple of N and $B/N + 1$ denotes the number of FDLs used in each optical output buffer. The third block comprises M FDL arrays, one for each of the M output ports. Each FDL array consists of $B/N + 1$ delay lines of different lengths. Specifically, each array is built of B/N FDLs whose lengths differ by the transmission time of a packet, T, as well as one additional fiber with an infinitesimal small length. By using TWCs the packet loss performance of OPS nodes is improved, especially for an increasing number of wavelengths. Furthermore, by exploiting the wavelength dimension by means of TWCs, the required number of FDLs in OPS nodes can be reduced. In fact, the flexibility gained by using TWCs allows for realizing OPS nodes without any FDLs, resulting in bufferless OPS node and network architectures.

Unified Contention Resolution

Optical buffering and wavelength conversion are effective approaches to resolve contention in OPS networks. Both techniques can be applied jointly or separately and have their respective merits and shortcomings. Unlike electrical RAM, optical buffers realized by FDLs offer only fixed and finite amounts of delay. To realize optical buffers of large capacity, OPS nodes are needed to deploy a large number of FDLs. Using FDLs not only deteriorates the optical signal quality but also increases delay and may cause packet reordering. On the other hand, wavelength conversion does not introduce any significant delay increase and avoids packet reordering. However, optical buffering and wavelength conversion have one thing in common that they require additional hardware at the OPS node. Optical buffering and wavelength conversion require the OPS node to be equipped with FDLs and TWCs, respectively.

Another approach to resolve contention in OPS networks without requiring hardware upgrades at any OPS node is deflection routing, which can be easily done in software. However, with deflection routing, optical packets may arrive out of order. Furthermore, its effectiveness strongly depends on the given network topology and traffic condition. To prevent deflected packets from looping endlessly throughout the OPS network, certain mechanisms must be deployed, for example, a maximum-hop count field may be used in the packet header, which is decremented at each intermediate OPS node such that the packet is discarded when the value of the field equals zero after the packet has traveled the maximum number of hops.

A unified contention resolution across wavelength, time, and space domains for mesh network topologies was examined in Ref. [24]. The proposed unified contention resolution explored various combinations of all three dimensions in terms of throughput, packet loss, and average hop distance. The obtained results indicate that wavelength conversion is the preferred technique to resolve contention in OPS networks, especially for an increasing number of wavelengths and under heavy traffic loads. Importantly, the study showed that deflection routing can be a good approach to resolve contention in OPS networks with high-connectivity topologies, but it is less effective in low-connectivity topologies. In general, deploying only deflection routing does not lead to significant performance improvements, and it was shown to be the least effective approach to resolve contention in OPS networks. However, deflection routing combined with wavelength conversion and possibly also optical buffering is a powerful approach to build high-performance OPS networks. Among all the considered combinational schemes, wavelength conversion, combined with carefully designed optical buffering and deflection routing at selected OPS nodes, appears to be the best approach to resolve contention in OPS networks.

PACKET OPTICAL TRANSPORT NETWORK

In their migration toward next-generation networks (NGNs), today's carriers desire to integrate the latest packet networking technologies such as connection-oriented Ethernet and multiprotocol label switching (MPLS) with installed optical network technologies such as WDM, ROADM, and OXC, giving rise to converged *packet optical transport networks* (P-OTNs). P-OTNs aim at providing a smooth migration platform toward an all-packet infrastructure and addressing carriers' drive to investment protection and lower CAPEX and OPEX by capitalizing on low-cost, layer-2 switching functionalities.

Even though IP/MPLS routers lead new installations, there exists a huge base of legacy synchronous optical network (SONET)/synchronous digital hierarchy (SDH) circuit-switched TDM equipment dominating the installed infrastructure of today's backbone networks. As a result of ITU-T's global standards initiative (GSI), current widely deployed circuit-switching SONET/SDH networks will evolve into NGNs whose transfer is based on packets instead of circuits in order to converge and optimize their operation and meet the increasing demand for new multimedia services and mobility.

Ethernet can greatly reduce the complexity and cost associated with the large scale and broad scope of carriers' circuit-switching networks by being a cost-effective and less complex replacement for SONET/SDH and taking advantage of Ethernet's low-cost points and simplicity. Ethernet is oriented toward simple but highly cost-effective and reliable packet switching. It has been a great success story as the packet-switching technology of choice in the vast majority of today's enterprise and local area networks (LANs).

The challenge is to develop traditional Ethernet in such a way that it provides the equivalent functionalities of existent SONET/SDH-based carrier networks without losing the low cost and simplicity that make Ethernet attractive in the first place, leading to *carrier Ethernet* [25]. Carrier Ethernet allows service providers to inexpensively interconnect end users for (virtual) private line and IP/MPLS services and has become the dominant backhaul interface for mobile services with the pending deployment of fourth-generation (4G) technology. Given that the volume of Ethernet traffic is growing at unprecedented rates, carrier Ethernet infrastructures and services become increasingly vital. This is witnessed by the fact that retail enterprise Ethernet ports are projected to grow at a 40% compound annual growth rate [26] between 2007 and 2012 and that more than 75% of service providers have a strategy of using carrier Ethernet instead of SONET/SDH for accessing and collecting customer traffic [27].

Toward the realization of carrier Ethernet networks, traditional Ethernet bridges/switches must be gradually enhanced with advanced capabilities and forwarding models, while at the same time operating at ever increasing line rates. The latest carrier Ethernet technology called *provider backbone bridge-traffic engineered* (PBB-TE) was recently ratified in IEEE standard 802.1Qay in June 2009. PBB-TE enhances carrier Ethernet networks with the support of traffic- engineered, point-to-point, and point-to-multipoint connections called *Ethernet switched paths* (ESPs). PBB-TE preserves the connectionless Ethernet behavior and adds a connection-oriented forwarding mode to current Ethernet networks by simply turning off traditional flooding and learning techniques. In doing so, the existing Ethernet control protocol (spanning tree protocol [STP] and its derivatives) and all its associated constraints and problems, for example, slow convergence time and lack of load balancing, disappear. PBB-TE enhances carrier Ethernet networks with traffic engineering, deterministic QoS, and support for protection switching at Ethernet cost points and virtually no learning curve. PBB-TE allows service providers to maximize network utilization and hence minimize the cost per bit carried and provides an ideal platform to emulate revenue-generating voice services.

Ethernet Frame

Carrier Ethernet involves various standardization efforts that aim at equipping Ethernet with transport features such as wide area scalability, fast recovery from network failures, advanced traffic engineering and end-to-end QoS, as well as OAM capabilities [28]. The evolution of traditional Ethernet switches into carrier Ethernet switches has involved several steps. The step-by-step migration toward carrier Ethernet implied a number of evolutionary modifications of the original IEEE 802.1 Ethernet frame format, depicted in Figure 34.15a. The Ethernet frame evolution was done by means of *tagging, stacking,* and *encapsulation.*

- *Tagging:* Tagging was initially used to realize customer virtual LANs (VLANs) and was specified in IEEE 802.1Q in 1998. As shown in Figure 34.15b, VLAN switches add the so-called *Q-tag* to Ethernet frames at the ingress node and remove it at the egress node of a VLAN switching network. Beside priority bits, the Q-tag contains a 12 bit VLAN identifier (VID) field, which is used to uniquely identify customer VLANs. A VLAN switch filters incoming Ethernet frames based on their VID and uses a separate forwarding table for each single VID or each subset of VIDs.

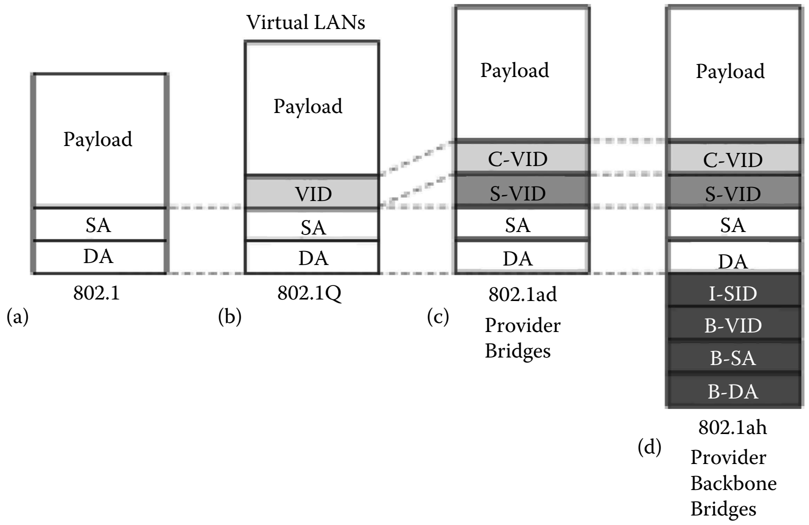

FIG. 34.15

Ethernet frame evolution. SA, source MAC address; DA, destination MAC address; VID, VLAN ID; C-VID, customer VID; S-VID, service VID; I-SID, service ID; B-VID, backbone VID; B-DA, backbone DA; B-SA, backbone SA. (After Allan, D., Bragg, N., McGuire, A., and Reid, A., Ethernet as carrier transport infrastructure, IEEE Communications Magazine, 44(2), 134, February 2006 © 2006 IEEE.)

- *Stacking:* In an effort to increase the VLAN space, *provider bridges* (PBs), specified in IEEE 802.1ad in 2005, separate the customer VLAN space from that of service providers through VLAN stacking of two different Q-tags. As shown in Figure 34.15c, the first Q-tag contains the service VID (S-VID) used by service provider PBs for forwarding, whereas the second Q-tag contains the original customer VID (C-VID). The stacking of the two Q-tags is often referred to as *Q-in-Q*.
- *Encapsulation:* To isolate the customer and provider domains from each other and thereby significantly improve the scalability of carrier Ethernet networks, *provider backbone bridges* (PBBs), specified in IEEE 802.1ah in 2008, encapsulate each customer frame within a provider frame. As shown in Figure 34.15d, PBBs evolve the Ethernet frame by adding an additional MAC header, also known as *MAC-in-MAC* encapsulation. PBBs prepend their own backbone MAC source address (B-SA) and destination address (B-DA) as well as a separate backbone VID (B-VID) and a new 24 bit field called service identifier (I-SID). PBBs significantly improve network scalability by using two separate tags, one for VLAN identification (B-VID) and one for service differentiation (I-SID). By using a separate service identifier field, PBBs allow for supporting up to 16 million service instances independently from the fields B-VID and B-DA, which are used for forwarding. As a result, PBBs are able to separate their forwarding decisions from the identification of service instances. PBBs make a clear separation between customer and service provider MAC domains and provide transparent Ethernet tunnels for customer Ethernet frames, resulting in improved security, stability, and a reduced burden on the PBB forwarding tables.

The forwarding tables of VLAN switches, PBs, and PBBs are populated through traditional flooding and MAC address learning techniques. Note that the aforementioned evolutionary carrier Ethernet upgrades affect only the header of Ethernet frames, while leaving the payload unchanged for backward compatibility. Thus, carrier Ethernet and traditional Ethernet frames alike can carry payload up to a maximum transmission unit (MTU) of 1500 bytes.

Recently, various proposals have been considered to add connection-oriented forwarding to carrier Ethernet. The three main approaches are the above mentioned PBB-TE; transport MPLS (T-MPLS), which is in the process of being superseded by the jointly developed IETF/ITU-T MPLS transport profile (MPLS-TP); and VLAN cross-connect (VXC), also known as provider VLAN transport (PVT) [29]. All of them still suffer from a number of shortcomings, for example, lack of support for multipoint-to-multipoint services and an adequate control plane. According to Sofia [29], however, PBB-TE leads the run in terms of backward compatibility and Ethernet services being defined by the Metro Ethernet Forum (MEF), which is a global industry alliance comprising more than 150 organizations.

Provider Backbone Bridge-Traffic Engineering

PBB-TE reserves a range of B-VIDs to identify explicitly routed ESPs. An ESP is point-to-point or point-to-multipoint. Each ESP is uniquely identified by a separate tuple <B-DA, B-SA, B-VID>. An ESP is bidirectional congruent, that is, it traverses the same nodes, ports, and links in both directions. A PBB-TE network contains two different bridge types: backbone edge bridges (BEBs) and backbone core bridges (BCBs). BEBs are located at the ingress and egress of the network and perform encapsulation and decapsulation of customer frames, respectively. In the core of a PBB-TE network, BCBs forward encapsulated frames based on their 60 bit <B-DA, B-VID> header information, which does not change along the corresponding ESP (i.e., no label swapping). For a given B-DA, a separate B-VID must be assigned to each ESP. The allocated B-VIDs are tied to the B-DA of the egress BEB and thus do not necessarily have to be globally unique. The forwarding table information at BCBs is provided by an external management or control plane instead of traditional flooding and learning. Multiple ESPs can be configured between a pair of BEBs for traffic engineering, load balancing, or protection.

The MEF defines a user network interface (UNI) and three different types of Ethernet virtual connection (EVC), which can be deployed in PBB-TE networks [30]. The MEF specifications associate several service attributes to each EVC and its UNIs. For instance, the *bandwidth profile* specifies the traffic characteristics of an EVC in terms of committed information rate, committed burst size, excess information rate, excess burst size, and color mode. The color mode indicates whether frames are marked (i.e., green, yellow, or red) at the ingress UNI for traffic policing.

CONCLUSIONS

While many optical technologies, for example, ROADM, OXC, WDM links, or PON, are successfully deployed in today's communications networks, some of the aforementioned switching techniques continue to face major roadblocks toward commercial adoption. Despite years of research, it remains to be seen whether OBS and OPS eventually become a reality given the limitations of current optics and photonics and the lack of techno-economic justification. On the other hand, optical Ethernet seems to be a promising technology that helps simplify the operation of fiber-optic networks and reduce their cost and complexity significantly. Despite recent progress, a number of open issues exist for recently standardized carrier Ethernet networks. Among others, an important feature of future network equipment is energy consumption in relation to network operation as well as heat dissipation. Lowering energy consumption results in significant cost savings and gives service providers a competitive edge. Furthermore, in today's Internet, voice and video applications are exploding with an ever-increasing number of users streaming multimedia content to their PCs and wireless handhelds, for example, iPhones, BlackBerrys, and peer-to-peer (P2P) services consuming more and more bandwidth. Novel node architectures with advanced packet switching capabilities and new forwarding models are needed to handle multimedia streams in a more resource-efficient and less power-consuming fashion.

References

1. Maier M., *Optical Switching Networks*, Cambridge, U.K.: Cambridge University Press, February 2008.
2. Ramaswami R., Optical networking technologies: What worked and what didn't, *IEEE Communications Magazine*, 44(9), 132–139, September 2006.
3. Koonen T., Fiber to the home/fiber to the premises: What, where, and when, *Proceedings of the IEEE*, 94(5), 911–934, May 2006.
4. Maier M. and Reisslein M., Trends in optical-switching techniques: A short survey, *IEEE Network*, 22(6), 42–47, November/December 2008.
5. Sengupta S., Kumar V., and Saha D., Switched optical backbone for cost-effective scalable core IP networks, *IEEE Communications Magazine*, 41(6), 60–70, June 2003.
6. Cao X., Anand V., and Qiao C., Waveband switching in optical networks, *IEEE Communications Magazine*, 41(4), 105–112, April 2003.
7. Cao X., Anand V., and Qiao C., Multi-layer versus single-layer cross-connect architectures for waveband switching, in *Proceedings of the IEEE INFOCOM*, Hong Kong, China, March 7–11, 2004, Vol. 3, pp. 1830–1840, 2004.
8. Li M., Yao W., and Ramamurthy B., Same-destination-intermediate grouping vs. end-to-end grouping for waveband switching in WDM mesh networks, in *Proceedings of the IEEE International Conference on Communications (ICC)*, Seoul, South Korea, May 23–28, 2005, Vol. 3, pp. 1807–1812, 2005.
9. Yao S., Mukherjee B., Yoo S. J. B., and Dixit S., A unified study of contention-resolution schemes in optical packet-switched networks, *IEEE/OSA Journal of Lightwave Technology*, 21(3), 672–683, March 2003.
10. Yao S. and Mukherjee B., Design of hybrid waveband-switched networks with OEO traffic grooming, in *Proceedings of the Optical Fiber Communication (OFC)*, Atlanta, GA, Vol. 1, pp. 357–358, 2003.
11. Chlamtac I., Fumagalli A., and Wedzinga G., Slot routing as a solution for optically transparent scalable WDM wide area networks, *Photonic Network Communications*, 1(1), 9–21, June 1999.
12. Elek V., Fumagalli A., and Wedzinga G., Photonic slot routing: A cost-effective approach to designing all-optical access and metro networks, *IEEE Communications Magazine*, 39(11), 164–172, November 2001.
13. Zang H., Jue J. P., and Mukherjee B., Capacity allocation and contention resolution in a photonic slot routing all-optical WDM mesh network, *IEEE/OSA Journal of Lightwave Technology*, 18(12), 1728–1741, December 2000.
14. Chan V. W. S., Hall K. L., Modiano E., and Rauschenbach K. A., Architectures and technologies for high-speed optical data networks, *IEEE/OSA Journal of Lightwave Technology*, 16(12), 2146–2168, December 1998.
15. Froberg N. M., Henion S. R., Rao H. G., Hazzard, B. K., Parikh, S., Romkey, B. R., and Kuznetsov, M., The NGI ONRAMP test bed: Reconfigurable WDM technology for next generation regional access networks, *IEEE/OSA Journal of Lightwave Technology*, 18(12), 1697–1708, December 2000.
16. Ganguly B. and Modiano E., Distributed algorithms and architectures for optical flow switching in WDM networks, in *Proceedings of the IEEE Symposium on Computers and Communications (ISCC)*, Antibes, France, July 4–6, 2000, pp. 134–139, 2000.
17. Battestilli T. and Perros H., An introduction to optical burst switching, *IEEE Communications Magazine*, 41(8), S10–S15, August 2003.
18. Verma S., Chaskar H., and Ravikanth R., Optical burst switching: A viable solution for terabit IP backbone, *IEEE Network*, 14(6), 48–53, November/December 2000.
19. Hunter D. K. and Andonovic I., Approaches to optical internet packet switching, *IEEE Communications Magazine*, 38(9), 116–122, September 2000.
20. Chang G.-K., Yu J., Yeo Y.-K., Chowdhury A., and Jia Z., Enabling technologies for next-generation optical packet-switching networks, *Proceedings of the IEEE*, 94(5), 892–910, May 2006.
21. El-Bawab T. S. and Shin J.-D., Optical packet switching in core networks: Between vision and reality, *IEEE Communications Magazine*, 40(9), 60–65, September 2002.
22. Hunter D. K., Chia M. C., and Andonovic I., Buffering in optical packet switches, *IEEE/OSA Journal of Lightwave Technology*, 16(12), 2081–2094, December 1998.
23. Danielsen S. L., Hansen P. B., and Stubkjaer K. E., Wavelength conversion in optical packet switching, *IEEE/OSA Journal of Lightwave Technology*, 16(12), 2095–2108, December 1998.
24. Yao S., Ou C., and Mukherjee B., Design of hybrid optical networks with waveband and electrical TDM switching, in *Proceedings of the IEEE GLOBECOM*, San Francisco, CA, Vol. 5, pp. 2803–2808, 2003.
25. Reid A., Willis P., Hawkins I., and Bilton C., Carrier Ethernet, *IEEE Communications Magazine*, 46(9), 96–103, September 2008.
26. Reddy P. and Lisle S., Ethernet aggregation and transport infrastructure OAM and protection issues, *IEEE Communications Magazine*, 47(2), 152–159, February 2009.
27. Infonetics Research, *Service Provider Plans for Metro Optical and Ethernet: North America, Europe, and Asia Pacific 2007,* September 2007.
28. Allan D., Bragg N., McGuire A., and Reid A., Ethernet as carrier transport infrastructure, *IEEE Communications Magazine*, 44(2), 134–140, February 2006.
29. Sofia R. C., A survey of advanced Ethernet forwarding approaches, *IEEE Communications Surveys and Tutorials*, 11(1), 92–115, First Quarter 2009.
30. Fouli K. and Maier M., The road to carrier-grade Ethernet, *IEEE Communications Magazine*, 47(3), S30–S39, March 2009.